

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-242881

(43)Date of publication of application : 07.09.2001

(51)Int.Cl.

G10L 13/04

G06F 3/16

G10L 13/08

(21)Application number : 2000-052857

(71)Applicant : NIPPON TELEGR & TELEPH CORP <NTT>

(22)Date of filing : 29.02.2000

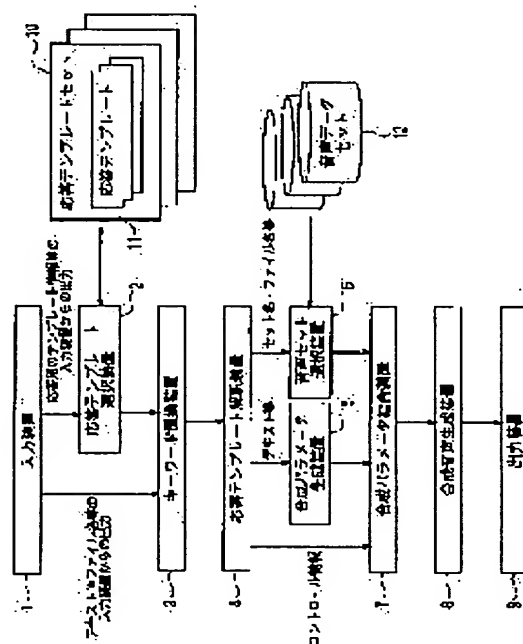
(72)Inventor : SHINOZAKI TASUKU
ABE MASANOBU

(54) METHOD OF VOICE SYNTHESES AND APPARATUS THEREOF

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a method and an apparatus for voice synthesis in which natural and smooth voice having a variety of feeling expressions is synthesized.

SOLUTION: A response template 11 is selected from a response template set 10. Voice synthesis parameters are generated by a synthesis parameter generator 5 and a voice set selector 6, based on the description of the template 11 and input information from an inputting device 1. Then, the parameters are combined based on the description of the template and voice is generated by a synthesized voice generator 8, based on the combined voice synthesis parameters.



LEGAL STATUS

[Date of request for examination]

16.11.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

3578961

[Date of registration]

23.07.2004

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

THIS PAGE BLANK (USPTO)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-242881

(P 2 0 0 1 - 2 4 2 8 8 1 A)

(43) 公開日 平成13年9月7日(2001.9.7)

(51) Int. Cl. ⁷	識別記号	F I	テマコード (参考)
G10L 13/04		G06F 3/16	330 K 5D045
G06F 3/16	330	G10L 5/02	J 9A001
G10L 13/08		3/00	H

審査請求 未請求 請求項の数 6 O L (全 8 頁)

(21) 出願番号 特願2000-52857(P 2000-52857)

(22) 出願日 平成12年2月29日(2000.2.29)

(71) 出願人 000004226

日本電信電話株式会社

東京都千代田区大手町二丁目3番1号

(72) 発明者 篠崎 翼

東京都千代田区大手町二丁目3番1号 日

本電信電話株式会社内

(72) 発明者 阿部 匡伸

東京都千代田区大手町二丁目3番1号 日

本電信電話株式会社内

(74) 代理人 100069981

弁理士 吉田 精孝

Fターム(参考) 5D045 AA08 AA09

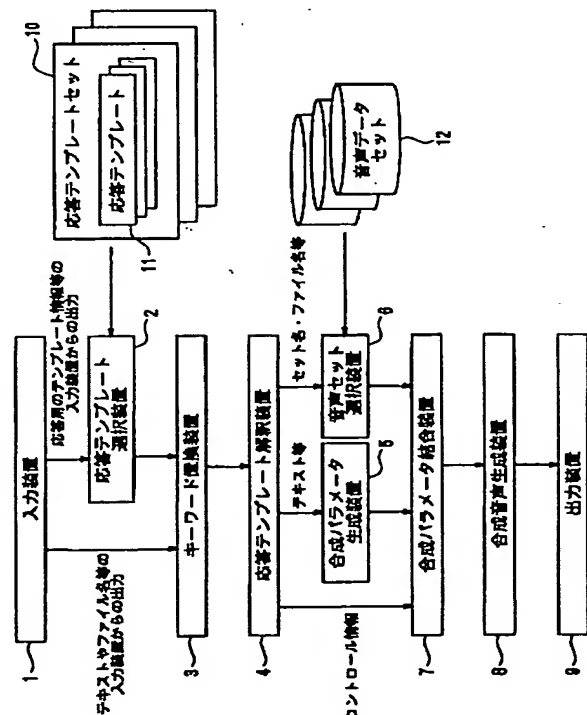
9A001 DD11 HH18

(54) 【発明の名称】 音声合成方法及び装置

(57) 【要約】

【課題】 感情表現など多様な表情をもった自然で滑らかな音声进行を合成できる音声合成方法及び装置を提供する。

【解決手段】 応答テンプレートセット10から応答テンプレート11を選択し、該応答テンプレート11の記述及び入力装置11からの入力情報に基づき合成パラメータ生成装置5及び音声セット選択装置6で音声合成用パラメータを作成し、該音声合成用パラメータを前記応答テンプレートの記述に基づき結合し、結合した音声合成用パラメータに基づき合成音声生成装置8で音声を生成する。



【特許請求の範囲】

【請求項 1】 音声合成用パラメータを作成するパラメータ作成手段と、音声合成用パラメータに基づき音声合成する音声合成手段とを備え、前記パラメータ作成手段により作成した音声合成用パラメータに基づき音声合成する音声合成方法において、

前記パラメータ作成手段において音声合成用パラメータの作成に用いるパラメータ作成情報及び作成された音声合成用パラメータの加工に用いるパラメータ加工情報を含む音声合成テンプレートを予め記憶手段に複数記憶し、

該記憶手段から音声合成テンプレートを選択し、選択した音声合成テンプレートのパラメータ作成情報に基づき前記パラメータ作成手段により音声合成用パラメータを作成し、

選択した音声合成テンプレートのパラメータ加工情報に基づき、前記パラメータ作成手段で作成した音声合成用パラメータを加工し、

加工した音声合成用パラメータに基づき前記音声合成手段により音声合成することを特徴とする音声合成方法。

【請求項 2】 前記パラメータ作成手段を複数設けるとともに、選択した音声合成テンプレートのパラメータ加工情報に基づき前記複数のパラメータ作成手段によりそれぞれ作成された複数の音声合成用パラメータを結合することを特徴とする請求項 1 記載の音声合成方法。

【請求項 3】 前記複数のパラメータ作成手段のうち少なくとも一つはパラメータ作成情報に含まれるテキストに基づき音声合成用パラメータを作成するようにし、

前記音声合成テンプレートのパラメータ作成情報にテキストを置換するための置換情報を含ませ、

選択した音声合成テンプレートのパラメータ作成情報に前記置換情報が含まれている場合には、該置換情報に基づき入力手段から入力されたテキストでパラメータ作成情報の一部を置換し、

前記パラメータ作成手段は置換したパラメータ作成情報に基づき音声合成用パラメータを作成することを特徴とする請求項 2 記載の音声合成方法。

【請求項 4】 音声合成用パラメータに基づき音声合成する音声合成手段を備えた音声合成装置において、

音声合成用パラメータの作成に用いるパラメータ作成情報及び作成された音声合成用パラメータの加工に用いるパラメータ加工情報を含む音声合成テンプレートを予め複数記憶した記憶手段と、

記憶手段から音声合成テンプレートを選択するテンプレート選択手段と、

テンプレート選択手段により選択された音声合成テンプレートのパラメータ作成情報に基づき音声合成用パラメータを作成するパラメータ作成手段と、

テンプレート選択手段により選択された音声合成テン

プレートのパラメータ加工情報に基づき、前記パラメータ作成手段により作成された音声合成用パラメータを加工するパラメータ加工手段とを備え、

前記音声合成手段は、パラメータ加工手段により加工された音声合成用パラメータに基づき音声合成すること

を特徴とする音声合成装置。

【請求項 5】 前記パラメータ作成手段を複数備えるとともに、

前記パラメータ加工手段は、前記テンプレート選択手段により選択された音声合成テンプレートのパラメータ加工情報に基づき前記複数のパラメータ作成手段によりそれぞれ作成された複数の音声合成用パラメータを結合することを特徴とする請求項 4 記載の音声合成装置。

【請求項 6】 前記記憶手段に記憶されている音声合成テンプレートのパラメータ作成情報はその一部をテキストで置換するための置換情報を含み、

テキストを入力する入力手段と、

前記テンプレート選択手段により選択された音声合成テンプレートのパラメータ作成情報に前記置換情報が含まれている場合には、該置換情報に基づき入力手段から入力されたテキストでパラメータ作成情報の一部を置換する置換手段とを備え、

前記複数のパラメータ作成手段のうち少なくとも一つは、前記置換手段により置換されたパラメータ作成情報に基づき音声合成用パラメータを作成することを特徴とする請求項 5 記載の音声合成装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、人—システムのインタフェース（マン・マシン・インタフェース）などに用いられる音声合成方法及び装置に関する。

【0002】

【従来の技術】 音声による応答は、電話による応答装置や金融機関の C D（現金自動支払装置）、A T M（現金自動預入支払装置）、各種自動販売機、電卓、音声を使ったおもちゃ、コンピュータ、音声対話システム、自動受付システム、C A I（コンピュータ支援教育）システムといった幅広い領域で利用されている。ここで用いられている音声としては、予め録音した自然音声、予め録音した音声をつなぎ合わせた音声（録音編集合成音声）、日本語文章を規則に従って変換した音声（規則合成音声）、音韻記号と F。（基本周波数）・継続時間・パワーといった韻律パラメータを操作して合成した音声（パラメータ合成音声）、波形編集合成音声などが挙げられる。

【0003】

【発明が解決しようとする課題】 従来の技術では、上記各種システムにおいて該システムを利用する人間（ユーザ）に対して上述した合成音声による応答を行っていた。ここで、応答音声を生成する場合、例えば予め作成

したパラメータ合成音声と規則合成音声（例えば任意性の高い名前や事柄等をテキスト等から合成した音声）を組み合わせる使用する場合には、パラメータ合成音声と規則合成音声のつなぎ目で不自然さが目立つ等、システムの応答として自然で滑らかな音声応答を行うことが困難であった。また、不自然さを避けるために応答音声の全てを予め用意すると、人名など場合によってはその組み合わせの余りの多さが問題になったり、逆に、全てを規則合成音声によってテキスト等から生成すると、機械的とか単調な応答になってしまう点が問題となっていた。

【0004】このため、従来の合成音声による音声応答方法や装置では、応答音声の接続が自然で滑らかである多様な応答音声システムに対して返すことができなかった。したがって、ユーザに対して機械的であるとか単調な印象・違和感を与えるため、システムを使う気にさせるとか、使って楽しいといったいわゆるユーザフレンドリネスをユーザに与えることが困難であった。

【0005】本発明は、上記事情に鑑みてなされたものであり、その目的とするところは、感情表現など多様な表情をもった自然で滑らかな音声を合成できる音声合成方法及び装置を提供することにある。

【0006】

【課題を解決するための手段】上記目的を達成するために、請求項1の発明では、音声合成用パラメータを作成するパラメータ作成手段と、音声合成用パラメータに基づき音声を合成する音声合成手段とを備え、前記パラメータ作成手段により作成した音声合成用パラメータに基づき音声を合成する音声合成方法において、前記パラメータ作成手段において音声合成用パラメータの作成に用いるパラメータ作成情報及び作成された音声合成用パラメータの加工に用いるパラメータ加工情報を含む音声合成テンプレートを予め記憶手段に複数記憶し、該記憶手段から音声合成テンプレートを選択し、選択した音声合成テンプレートのパラメータ作成情報に基づき前記パラメータ作成手段により音声合成用パラメータを作成し、選択した音声合成テンプレートのパラメータ加工情報に基づき、前記パラメータ作成手段で作成した音声合成用パラメータを加工し、加工した音声合成用パラメータに基づき前記音声合成手段により音声を合成することを特徴とするものを提案する。

【0007】本発明によれば、音声合成テンプレートを選択することにより、選択した音声合成テンプレートのパラメータ作成情報に基づき音声合成用パラメータが作成され、さらに該音声合成テンプレートのパラメータ加工情報に基づき音声合成用パラメータが加工される。そして、加工された音声合成用パラメータに基づき音声が合成される。これにより、予め音声合成テンプレートのパラメータ作成情報及びパラメータ加工情報を予め作成しておけば、パラメータ作成手段により作成された音声

合成用パラメータに対して、例えば滑らかな音声となるような種々の加工を行うことができる。また、特にパラメータ作成手段として従来の種々の方式を用いることができるので容易且つ簡便な構造で多様な音声合成を行うことができる。

【0008】また、請求項2の発明では、請求項1記載の音声合成方法において、前記パラメータ作成手段を複数設けるとともに、選択した音声合成テンプレートのパラメータ加工情報に基づき前記複数のパラメータ作成手段によりそれぞれ作成された複数の音声合成用パラメータを結合することを特徴とするものを提案する。

【0009】本発明によれば、音声合成テンプレートのパラメータ作成情報に基づき、複数のパラメータ作成手段によりそれぞれ音声合成用パラメータが作成され、音声合成テンプレートのパラメータ加工情報に基づき、この複数の音声合成用パラメータが結合される。したがって、このパラメータ加工情報を複数の音声合成用パラメータが滑らかで自然に結合されるように記述すれば、異なるパラメータ作成手段で作成された音声合成用パラメータが滑らかで自然に結合される。これにより、滑らかで自然な音声を合成することができる。

【0010】さらに、請求項3の発明では、請求項2記載の音声合成方法において、前記複数のパラメータ作成手段のうち少なくとも一つはパラメータ作成情報に含まれるテキストに基づき音声合成用パラメータを作成するようにし、前記音声合成テンプレートのパラメータ作成情報にテキストを置換するための置換情報を含ませ、選択した音声合成テンプレートのパラメータ作成情報に前記置換情報が含まれている場合には、該置換情報に基づき入力手段から入力されたテキストでパラメータ作成情報の一部を置換し、前記パラメータ作成手段は置換したパラメータ作成情報に基づき音声合成用パラメータを作成することを特徴とするものを提案する。

【0011】本発明によれば、入力したテキストに基づき音声合成用パラメータが作成されるとともに、該音声合成用パラメータと他のパラメータ作成手段により作成された音声合成用パラメータとをパラメータ加工情報に基づき結合される。これにより、任意のテキストを含む文章を自然で滑らかな音声に合成することができる。

【0012】さらに、請求項4の発明では、音声合成用パラメータに基づき音声を合成する音声合成手段を備えた音声合成装置において、音声合成用パラメータの作成に用いるパラメータ作成情報及び作成された音声合成用パラメータの加工に用いるパラメータ加工情報を含む音声合成テンプレートを予め複数記憶した記憶手段と、記憶手段から音声合成テンプレートを選擇するテンプレート選擇手段と、テンプレート選擇手段により選擇された音声合成テンプレートのパラメータ作成情報に基づき音声合成用パラメータを作成するパラメータ作成手段と、テンプレート選擇手段により選擇された音声合成テンプレ

10

20

30

40

50

レートのパラメータ加工情報に基づき、前記パラメータ作成手段により作成された音声合成用パラメータを加工するパラメータ加工手段とを備え、前記音声合成手段は、パラメータ加工手段により加工された音声合成用パラメータに基づき音声合成することを特徴とするものを提案する。

【0013】さらに、請求項5の発明では、請求項4記載の音声合成装置において、前記パラメータ作成手段を複数備えとともに、前記パラメータ加工手段は、前記テンプレート選択手段により選択された音声合成テンプレートのパラメータ加工情報に基づき前記複数のパラメータ作成手段によりそれぞれ作成された複数の音声合成用パラメータを結合することを特徴とするものを提案する。

【0014】さらに、請求項6の発明では、請求項5記載の音声合成装置において、前記記憶手段に記憶されている音声合成テンプレートのパラメータ作成情報はその一部をテキストで置換するための置換情報を含み、テキストを入力する入力手段と、前記テンプレート選択手段により選択された音声合成テンプレートのパラメータ作成情報に前記置換情報が含まれている場合には、該置換情報に基づき入力手段から入力されたテキストでパラメータ作成情報の一部を置換する置換手段とを備え、前記複数のパラメータ作成手段のうち少なくとも一つは、前記置換手段により置換されたパラメータ作成情報に基づき音声合成用パラメータを作成することを特徴とするものを提案する。

【0015】これら、請求項4乃至6の発明に係る音声合成装置によれば、上記請求項1乃至3に記載した音声合成方法を実現することができる。

【0016】

【発明の実施の形態】本発明の一実施の形態に係る音声合成方法及び装置について図面を参照して説明する。本実施の形態では本発明の音声合成方法及び装置を外部からの入力に対して音声で応答を行う音声応答装置に適用した例を示す。図1は音声応答装置の構成図である。

【0017】図1において、1は音声応答装置への入力装置であり、例えばキーボード、マウス・タッチパネル等のポインティングデバイス、他の装置とのインタフェースデバイス等である。入力装置1では、音声応答装置への入力情報から、使用する応答テンプレートの選択基準となるテンプレート情報と、応答テンプレート中のキーワードを置換するためのテキストやファイル名等を入力する。

【0018】図1において、2は応答テンプレート選択装置である。この応答テンプレート選択装置2は、入力装置1からのテンプレート情報に基づいて、複数の応答テンプレートセット10から使用する応答テンプレートセット10の中から指定された応答テンプレート11を決定

する。ここで、応答テンプレートセット10は、例えば「挨拶」、「質問」、「謝罪」のように応答のカテゴリ別にセットを構成し、例えば応答テンプレートセット10が「謝罪」であれば、「ごめんなさい。」、「ごめんね。」、「ごめん。」のように応答テンプレートセット10毎に複数の応答テンプレート11を用意している。これにより、同じ内容の応答（同じカテゴリの内容）でも多様な応答をすることができる。

【0019】図1において、3はキーワード置換装置であり、応答テンプレート選択装置2の出力である選択された応答テンプレート11中のキーワード部分を、入力装置1からの出力であるキーワード置換用のテキストやファイル名等に置換する。例えば応答テンプレート11に「KeyWD__NAMEちゃん、こんばんは。」と記述している場合、入力装置1からの「KeyWD__NAME=”翼”」という置換情報に基づきキーワードの置換を行い、応答テンプレート11を「翼ちゃん、こんばんは。」と書き換える。このように、応答テンプレート11中にキーワードを記述し置換することで固定的な応答ではなく多様で柔軟な応答を返すことができる。

【0020】図1において、4は応答テンプレート解釈装置である。この応答テンプレート解釈装置4は、キーワード置換装置3の出力に基づいて応答テンプレート11中に記述された処理内容をテンプレートの先頭から解釈し、応答音声を作成するための処理を制御する。

【0021】図1において、5は合成パラメータ生成装置である。この合成パラメータ生成装置5は、応答テンプレート解釈装置4から出力されたキーワード置換済みの応答テンプレート11の記述に基づいて、応答テンプレート11中のテキスト及びアクセント記号付きテキストを解析し、応答テンプレート11中の合成コントロール情報の記述にしたがって、音声を合成するのに必要な音声合成用パラメータの一つである韻律情報と音韻記号を付加した合成パラメータを生成する。

【0022】図1において、6は音声セット選択装置である。この音声セット選択装置6は応答テンプレート11中の音声データのファイル名に該当する部分を音声データセット13から取得する。音声データセット12は、例えば波形データや合成パラメータ等の音声合成に必要な音声合成用パラメータが格納されたファイル群で構成される。

【0023】図1において、7は合成パラメータ結合装置である。この合成パラメータ結合装置7は、合成パラメータ生成装置5と音声セット選択装置6から出力された各種音声合成用パラメータ及びテンプレート解釈装置4の出力であるコントロール情報に基づいて、各々の合成パラメータ及び波形データを自動的に編集・結合等の加工を行う。

【0024】図1において、8は合成音声生成装置である。この合成音声生成装置8は、合成パラメータ結合装

置 7 によって 1 つにまとめられた音声合成用パラメータである合成パラメータ及び波形データに基づいて合成音声を生成する。合成音声生成装置 8 で生成された音声は、出力装置 9 により出力される。

【0025】図 2 は応答テンプレート 11 の記述例である。図 2 の例では、「やあこんにちは、つばさちゃん、元気だった？」という応答音声を生成するための応答テンプレートである。応答テンプレート 11 には合成するためのテキストやアクセント記号付きのテキスト及び合成パラメータや波形データのついたファイル名等（データ）のパラメータ作成情報が、それらの結合方法及び音声合成の際に指定する音声合成のパラメータ等の指定記述（コントロール情報）とともに記述されている。すなわち、応答テンプレート 11 には、合成パラメータ生成装置 5 及び音声セット選択装置 6 で音声合成用パラメータの作成に必要なパラメータ作成情報と、合成パラメータ生成装置 5 及び音声セット選択装置 6 の各出力を加工（滑らかな結合等）するのに必要な加工情報が記述されている。図 2 では、応答テンプレートは、括弧内に記述されたコントロール情報（パラメータ加工情報）と、これに続くデータ（パラメータ作成情報）の列で記述されている。なお、図 2 において、「#」はコメントを表す記号であり該記号以降の文章に機能的な意味はない。

【0026】図 2 の具体例では、「こんにちは」、「ちゃん」、「だった？」といった予め作成された合成パラメータが格納されたファイル（それぞれファイル名: First.Pac, Second.Pac, Third.Pac）が用意されており、「やあ」、「つばさ」というテキストと「ゲンキ[01]」というアクセント記号付きテキストと組み合わせることで音声応答を生成する。この例では、まず「やあ」について語尾の母音を 450ms に伸ばすと同時に最後のピッチを標準より 25Hz 低くする。次に、First.Pac（「こんにちは」）を「やあこんにちは」と一つの単語のように合成するために First.Pac のパラメータを修正する。次いで、「やあこんにちは」と音の高さに歪みが生じないように、「つばさ」の平均ピッチと語全体のピッチの傾を修正し語尾の後のポーズ長を 0ms にする。さらに、「やあこんにちは、つばさ」と Second.Pac（「ちゃん」）を一つの単語のように合成するために Second.Pac の合成パラメータを修正する。同様に第 1 モーラにアクセント核のある「ゲンキ[01]」と Third.Pac（「だった」）を一つの単語のように合成するために Third.Pac の合成パラメータを修正し、結合することで応答音声「やあこんにちは、つばさちゃん、元気だった？」を生成する。このように応答テンプレート 11 にそれぞれのデータの結合順序と結合方法等（コントロール情報）を細かく記述することで、応答音声の自然で滑らかな接続を実現できる。

【0027】図 3 乃至図 6 は結合方法等のコントロール

情報の一例を示したものである。コントロール情報には主に接頭と語尾、語全体についての指示がある。また、それぞれに、音の高さを決めるピッチと音の大きさを決めるパワー、発話の早さを決める話速についての指示がある。そして、これら指定した値の効果をどのように及ぼすかを定義した形状を指定する（図 4 参照）。この例では、リニアに効果を及ぼす Type1 と最後に急激に効果を及ぼす Type2、最初に急激に効果を及ぼす Type3 の 3 つの形状を示している。また、語頭の前及び語尾の後のポーズ長や 2 つ以上のデータを 1 つの語句として結合するためのコントロール情報、男声や女声などの話者、デフォルトの話速等の合成エンジンへの合成パラメータ等がある。

【0028】図 6 で示したコントロール情報の例を簡単に説明すると、E1 のコントロール情報は、語頭のピッチをデータの全体長の 30% の位置で急激に立ち上げ 300Hz にする。E2 のコントロール情報は、語尾のピッチをデータの全体長の 20% の位置からリニアに徐々に下げてデフォルトのピッチより 15Hz 低くする。E3 のコントロール情報はコントロール情報は語尾から数えて 2 モーラ目からピッチに急減な変化を与え語尾のピッチを 220Hz にする。E4 のコントロール情報は語尾から数えて最初の母音のピッチに急激な変化を与え語尾のピッチを 220Hz にする。

【0029】次に、この音声応答装置を用いたシステムについて図 7 を参照して説明する。ここでは、しりとりシステムを例にとって具体的に説明する。まず、しりとりシステムについて説明する。しりとりシステムは、ユーザとシステムとがしりとりを行うシステムである。ユーザは、まず自分の名前を入力し、「乗り物」や「食べ物」といったしりとりのカテゴリを選択する。システムはユーザが入力したしりとり単語を、システムが持っている辞書から検索し答えを返す。このようにしてシステムとユーザがしりとりを行う。

【0030】このシステムでは、入力装置 1 による入力される情報は、キーボードから入力されたユーザの文字列、しりとりシステムの辞書から検索された単語、ユーザの名前、しりとりのカテゴリである。

【0031】応答テンプレート選択装置 2 は、「出たしの言葉が違う」「辞書にない単語である」「前に使った単語である」といったしりとりシステムからの出力イベントに応じて、イベント毎に用意した応答テンプレートセットから応答テンプレートセット 10 を選択する。図 7 にイベント「出たしの言葉が違う」用の応答テンプレートセット 10 の例を示す。図 7 には 3 つの応答テンプレート 11a ~ 11c が記述されている。なお、図 7 において「REM」はコメント表す記号であり該記号以降の文章に機能的な意味はない。

【0032】このしりとりシステムでは、3 つの応答テンプレート 11a ~ 11c の中から 1 つをランダムに選







択して応答を返す。ここでは、応答テンプレート 11c を例にとって説明する。

【0033】 応答テンプレート 11c では、11_4.Pac (「なんです?」) の合成パラメータに続けて、「%U」で示されたキーワードをユーザが入力した単語「ライト」に置換し、単語「ライト」の語尾「ト」の母音の継続時間を 200ms に変更し、ピッチを現在の値から相対的に 200Hz 上昇させる。次に「%e」で示されたキーワードをユーザが入力した単語「ライト」の語尾の 1 モーラ「ト」で置換し、合成エンジンへのデフォルト話速を 16 に設定し、合成パラメータを得た後に「ト」の母音の継続時間長を 300ms に変更することで「トー」にする。このとき「トー」の後のポーズを 0ms にする。これを 3 回行う。そして、13_11_1.Pac (「あれ? あれあれ? と思ったけど、出だしが違うよ。」) と単純に結合した後、「%E」で示されたキーワードを、システムが前回出力した単語の最後の 1 モーラである「ナ」で置換し、6_2_2.Pac (「で始まるはずだよ。もう一回考え直し。」) と接続する。応答テンプレートの処理は、キーワード置換装置 3、テンプレート解釈装置 4、合成パラメータ生成装置 5、音声セット選択装置 6、合成パラメータ結合装置 7、そして合成音声生成装置 8 でそれぞれ行われ、応答音声「なんです? ライト? トー トー トー、あれ? あれあれ? と思ったけど、出だしが違うよ。ナで始まるはずだよ。もう一回考え直し。」が生成される。

【0034】 このように、予め韻律を調整して感情表現などを行った合成音声を部品とした音声セットを構築しておき、応答テンプレートセットで用いることで、多様な応答が行える。そして、応答テンプレートに合成パラメータのファイル名やテキスト、アクセント記号付きテキスト等の個々のデータの接続方法を細かく記述することで、それぞれの接続が自然で滑らかな応答音声を生成できる。また、応答テンプレート中にキーワードを記述した後で応答音声を生成するときに前記テキスト等で置換することにより、応答音声の一部を任意に変えることができ任意の応答を生成できる。これにより、システムに対して、機械的であるとか単調な印象や接続の不自然さによる違和感といったものを与えるような音声を提供す

【図 4】

形状の記述例

	Type1	Type2	Type3
上昇(+)			
下降(-)			

ることがない。したがって、ユーザに対して、システムを使って楽しいとか、システムを使う気になるといった、いわゆるユーザフレンドリなインタフェースを提供することができる。

【0035】

【発明の効果】 以上詳述したように、本発明によれば、音声合成テンプレートを選択することにより、選択した音声合成テンプレートのパラメータ作成情報に基づき音声合成用パラメータが作成され、さらに該音声合成テンプレートのパラメータ加工情報に基づき音声合成用パラメータが加工される。そして、加工された音声合成用パラメータに基づき音声合成される。これにより、予め音声合成テンプレートのパラメータ作成情報及びパラメータ加工情報を予め作成しておけば、パラメータ作成手段により作成された音声合成用パラメータに対して、例えば自然で滑らかな音声となるような種々の加工を行うことができる。また、特にパラメータ作成手段として従来の種々の方式を用いることができるので容易且つ簡便な構造で感情表現など多様な表情をもった音声合成を行うことができる。

【図面の簡単な説明】

【図 1】 音声応答装置の構成図

【図 2】 応答テンプレートの例を説明する図

【図 3】 コントロール情報の例を説明する図

【図 4】 コントロール情報における形状指定の例を説明する図

【図 5】 コントロール情報における各項目の例を説明する図

【図 6】 コントロール情報の例を説明する図

【図 7】 しりとりシステムにおける応答テンプレートセットの例を説明する図

【符号の説明】

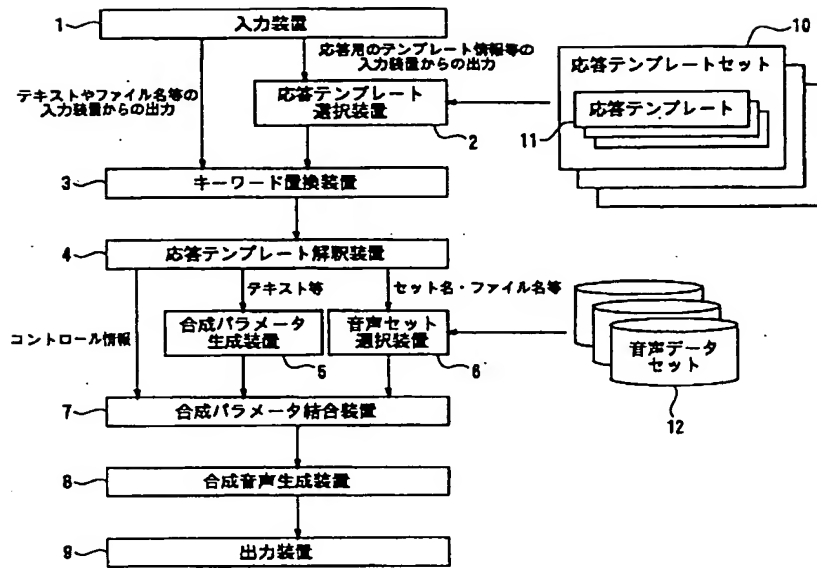
1…入力装置、2…応答テンプレート選択装置、3…キーワード置換装置、4…応答テンプレート解釈装置、5…合成パラメータ生成装置、6…音声セット選択装置、7…合成パラメータ結合装置、8…合成音声生成装置、9…出力装置、10…応答テンプレートセット、11…応答テンプレート、12…音声データセット

【図 6】

コントロール情報の具体例

E1	Pitch_Start=300Hz, Type2, 30%
E2	Pitch_End=15Hz, Type1, 20%
E3	Pitch_End=220Hz, Type3, 2more
E4	Pitch_End=220Hz, Type3, LastBowel

【図 1】



【図 2】

応答テンプレート 11

```

# First.Pac : 「こんにちは」
# Second.Pac : 「ちゃん、」
# Third.Pac : 「だった？」

(Duration_End=450ms,Type1,LastBowl;Pitch_End=25Hz,Type2,LastBowl)“やあ”
.(Connection=SameWord)First.Pac
.(Pitch_Average=Same;Duration_End=0)“つばさ”
.(Connection=SameWord)Second.Pac
.(GenKi[01])
.(Connection=SameWord)Third.Pac
# 出力 : 「やあこんにちは、つばさちゃん、元気だった？」

```

【図 5】

各項目の記述例

周波数	相対値指定又は絶対値指定(Hz等の単位)
パワー	相対値指定又は絶対値指定(dB等の単位)
位置	語頭又は語尾からの割合(%)による指定 語頭又は語尾からのモーラ数(mora)による指定 語頭又は語尾からの母音(Fast/LastBowl)による指定
時間長	絶対値指定(ms等の単位) 相対値指定(%)
話速	絶対値指定(mora/s等の単位) 相対値指定(%)

【図 3】

コントロール情報の一例

コントロール情報	記述
語頭のビッチの高さ	Pitch_Start=周波数,形状,位置
語尾のビッチの高さ	Pitch_End=周波数,形状,位置
語頭のパワーの大きさ	Power_Start=パワー,形状,位置
語尾のパワーの大きさ	Power_End=パワー,形状,位置
平均ビッチの高さ	Pitch_Average=周波数
平均パワーの大きさ	Power_Average=パワー
平均話速	Duration_Average=話速
全体話速	Duration_Average=時間長
語頭のポーズの長さ	Space_Start=時間長
語尾のポーズの長さ	Space=時間長
語頭の話速	Duration_Start=話速,形状,位置
語尾の話速	Duration_End=話速,形状,位置
結合形状の指定	Connection=SameWord(同一語句として)
その他	合成エンジンへの合成パラメータ (例えば、話者、デフォルト話速、平均ビッチ等)

【図 7】

応答テンプレートセット 10

